

УДК 371.26
ББК 74.00
Ш 65

О СЕРТИФИКАЦИИ КАЧЕСТВА ПЕДАГОГИЧЕСКИХ ТЕСТОВЫХ МАТЕРИАЛОВ

О. Ф. Шихова

Ключевые слова: тестирование, сертификация, педагогические тестовые материалы, содержательность, репрезентативность, латентность, когнитивность, валидность.

Резюме: В статье рассматриваются этапы и критерии сертификации качества педагогических тестовых материалов.

Совершенствование процесса тестирования с целью повышения его эффективности – одна из актуальных проблем российского образования. Ее появление связано с резким увеличением масштабов тестирования, введением государственных образовательных стандартов (ГОС), требующих модернизации систем контроля и оценки качества подготовки обучающихся.

С целью совершенствования аттестационных тестовых технологий Министерством образования Российской Федерации был издан приказ о сертификации качества педагогических тестовых материалов (ПТМ), предусматривающий, в том числе, и подготовку документов, регламентирующих саму процедуру сертификации. Опубликованы также Временное положение о сертификации качества ПТМ и методические указания по подготовке их к этой процедуре [5].

Согласно Временному положению, цель сертификации заключается в повышении качества образования за счет совершенствования педагогических тестовых материалов, которые применяются для оценки знаний обучающихся в учебных заведениях Российской Федерации. При этом последовательно решаются две задачи: выявление и фиксация достигнутого уровня качества сертифицируемых тестовых материалов и разработка рекомендаций, определяющих методику совершенствования ПТМ.

Процедура сертификации качества ПТМ осуществляется в два этапа. На первом производится экспертиза тестовых материалов, предусматривающая не только оценку качества их содержания и статистических характеристик, но и оценку правильности форм заданий в ПТМ. На втором этапе педагогическим тестовым материалам, прошедшим экспертизу и удовлетворяющим утвержденным требованиям к их качеству, выдается сертификат качества, который дает право на использование ПТМ для оценки знаний обучающихся.

При подготовке педагогических тестовых материалов к процедуре сертификации авторы должны самостоятельно оценить качество их содержания,

проверить соответствие заданий требованиям тестовой формы и выявить статистические характеристики ПТМ. Причем при проведении их содержательной экспертизы рекомендуется (согласно методическим указаниям [5]) использовать следующие критерии:

- полнота отображения материала образовательной программы при отборе содержания ПТМ;
- правильность пропорций, выбранных при отображении разделов предмета (содержательных линий);
- полнота охвата требований государственных образовательных стандартов в ПТМ;
- соответствие содержания заданий знаниям, умениям и навыкам, запланированным для проверки в спецификации ПТМ;
- значимость содержания каждого задания ПТМ для целей проверки.

Однако методика расчета перечисленных критериев в вышеназванных документах не приводится и, следовательно, на практике они будут определяться в основном интуитивно, без учета факторов, формирующих качество ПТМ. Между тем перечисленные критерии так или иначе характеризуют валидность тестовых материалов, то есть степень соответствия структуры и содержания ПТМ задачам и целям педагогического контроля. Исследователи выделяют множество различных видов валидности (содержательная, соответствия, прогностическая и т. д.), но, по сути, важнейшими для педагогического тестирования являются два вида: по содержанию и по критерию [4].

Содержательная валидность определяет «репрезентативность содержания дидактического теста выборке измеряемой области навыков и знаний учебного предмета» [3, с. 21]. При определении содержательной валидности необходимо установить степень соответствия заданий теста цели диагностики, в качестве которой могут выступать требования ГОСов; требования к профессиональной подготовке; установленные уровни обученности и т. п.

Критериальная валидность (или валидность соответствия) показывает соответствие между результатами тестирования и другими формами контроля, но с использованием в них тех же заданий, что и в тесте.

Количественной мерой валидности служит обычно коэффициент корреляции Пирсона или коэффициент связи ϕ (фи), данные для расчета которых получаются путем оценки группой опытных независимых экспертов предъявляемых им ПТМ по тому или иному критерию.

Однако, как показывает опыт, необходимых условий для организации педагогической экспертизы ПТМ и математико-статистической обработки ее результатов в большинстве вузов нет. Кроме того, большая часть преподавателей вообще не владеет основами тестологии. Поэтому процедура определения

валидности, характеризуемой набором вышеназванных критериев, должна быть достаточно проста и понятна.

Предлагаемый подход [2, 8] отвечает этим условиям, он достаточно универсален и позволяет не только количественно выразить установленные критерии, но и оперативно определить их. В качестве основных предлагаются такие критерии, как «Содержательность» (S), «Репрезентативность структуры» (R), «Репрезентативность ГОС» (G), «Латентность» (L) и «Когнитивность» (K).

Критерий «Содержательность» определяет полноту отображения материала образовательной программы в содержании тестовых заданий и рассчитывается по формуле:

$$S = \frac{N_T}{N_P},$$

где N_P – число дескрипторов (базовых понятий, терминов, формул и т. п.) в тезаурусе учебной программы (учебный тезаурус – это множество дескрипторов и множество связей между ними [6]);

N_T – число дескрипторов в тезаурусе ПТМ.

Значение данного критерия можно существенно повысить за счет расширения тезауруса педагогических тестовых материалов путем создания интегрированных, многофункциональных тестовых заданий.

Критерий «Репрезентативность структуры» позволяет сопоставить количество тестовых заданий в ПТМ по отдельным разделам (или темам) учебного предмета с объемом (в часах) этих разделов (или тем), то есть характеризует правильность пропорций при отображении в тесте разделов учебного предмета. Рассчитывается данный критерий по формуле:

$$R = 1 - \left(\frac{V_k}{V_s} - \frac{L_k}{L_s} \right),$$

где V_s – объем (в часах) учебного предмета;

V_k – объем (в часах) k -го диагностируемого модуля (например, одного из разделов физики);

L_s – общее количество заданий в тесте,

L_k – количество заданий для диагностики k -го модуля.

Если $R > 1$, то количество тестовых заданий в ПТМ по данному разделу (или теме) завышено. Если $R < 1$, то количество заданий в тесте по данному разделу занижено. Таким образом, рассчитав критерий «Репрезентативность

структуры», можно провести соответствующую корректировку теста, стремясь приблизить значения этого критерия к единице.

Критерий «Репрезентативность ГОС» устанавливает соответствие содержания тестовых материалов требованиям государственного образовательного стандарта по изучаемой дисциплине с учетом модели специалиста, то есть характеризует полноту охвата требований ГОС в ПТМ. Рассчитывается этот критерий по формуле:

$$G = \frac{T_d}{T_G},$$

где T_G – количество требований к специалисту в ГОС (или дисциплинарных требований к студенту);

T_d – количество требований, диагностируемых с помощью ПТМ.

Критерий «Латентность» характеризует пригодность тестовых материалов для диагностики уровней обученности в соответствии с выбранной таксономической моделью обучения как способа описания диагностируемых целей обучения. Это может быть шестиуровневая модель Б. Блума («знание», «понимание», «применение», «анализ», «синтез», «оценка»), четырехуровневая – В. П. Беспалько («узнавание», «воспроизведение», «эвристический», «творческий»), трехуровневая – ЗУН («знания», «умения», «навыки») и другие. Термин «Латентность» введен в связи с тем, что уровень обученности относится к латентным (скрытым) параметрам, ответственным за фиксируемые в опыте тестовые результаты. Предлагаемый критерий позволяет определить соответствие содержания заданий в тесте – знаниям, умениям и навыкам, запланированным для проверки в спецификации ПТМ, и рассчитывается по формуле:

$$L = \frac{l_d}{l_m},$$

где l_m – общее количество уровней обученности в выбранной модели обучения;

l_d – количество уровней обученности, диагностируемое тестом.

Критерий «Когнитивность» позволяет выявить пригодность ПТМ для оценки сформированности у обучаемых определенных видов знаний по классификации Б. Блума – Р. Гагна – В. С. Аванесова [1]. Этот классификатор сформулирован исключительно для решения задач педагогического измерения и не образует пока полной классификационной системы, но его использование позволяет более аргументированно подойти к отбору педагогических тестовых материалов. ПТМ должны разрабатываться с учетом принципа «возрастающей

трудности» учебного материала и включать задания для проверки сформированности как простейших знаний (знание названий, имен, фактов, определений и т. п.), так и алгоритмических, системных, вероятностных, методологических и т. п. знаний.

Рассчитывается критерий «Когнитивность» по формуле:

$$K = \frac{k_d}{k_N},$$

где K_N – общее количество видов знаний в принятой модели обучения;

K_d – количество видов знаний, диагностируемых с помощью ПТМ.

Данный критерий может учитываться при моделировании адаптивных тестов, позволяющих соотносить уровень подготовленности тестируемых с трудностью тестовых заданий.

К дополнительным критериям можно отнести: «Адаптивность» (мера соответствия среднего логита трудности заданий среднему логиту уровня обученности в выборке); «Профильность» (характеризует распределение заданий в ПТМ по схеме «возрастающей трудности» с целью повышения дифференцирующей способности тестовых материалов); «Системность» (характеризует разнообразие применяемых форм тестовых заданий, типов дидактических тестов, видов тестирования и оптимальность количества заданий в ПТМ); «Верифицируемость» (повторяемость результатов диагностики в различных выборках, статистическая надежность результатов диагностики) и др. [2, 8, 7].

Совокупность предлагаемых критериев сводит к минимуму влияние субъективных факторов при отборе и сертификации педагогических тестовых материалов. Это свойство ПТМ называется «объективированностью».

Общая оценка объективированности (качества) ПТМ может быть произведена по формуле:

$$\eta = C_1 S + C_2 R + C_3 G + C_4 L + C_5 K,$$

где $C_1 + C_5$ – соответствующие весовые коэффициенты, то есть коэффициенты «важности» критериев 1–5; их сумма должна быть равна единице (условие нормировки: $\sum_{i=1}^5 C_i = 1$).

Критерии 1–5 и их весовые коэффициенты определяются методом групповых экспертных оценок (ГЭО) [6]. В отличие от обычных комиссий в методе ГЭО эксперты работают независимо друг от друга, процедура отбора и опроса экспертов, разработка анкет и обработка полученных экспертных оценок про-

изводятся по определенному алгоритму в соответствии с требованиями ГОСТ 23554.1–79 [9].

Критерии с незначительными весовыми коэффициентами могут быть исключены при окончательной оценке ПТМ.

Процедура объективизации ПТМ будет заключаться в увеличении численных значений критериев за счет корректировки тестов с учетом результатов педагогического эксперимента. Все это позволит подготовить тестовые материалы к сертификации и, тем самым, будет способствовать повышению эффективности педагогического контроля в любом образовательном учреждении.

Литература

1. Аванесов В. С. Композиция тестовых заданий. – 2 – е изд. – М.: Изд-во «Адепт», 1998. – 217 с.
2. Лебедева Е. Н. Методические рекомендации по сертификации дидактических тестов для диагностики уровня обученности студентов / Под ред. В. С. Черепанова. – Ижевск: Изд-во ИЖГСХА, 1998. – 37 с.
3. Мирошниченко А. А., Казаринов А. С., Кулышева А. Ю. Технология адаптационной валидности: Учебное пособие. – Глазов: Изд-во ГГПИ, 1999. – 62 с.
4. Нейман Ю. М., Хлебников В. А. Введение в теорию моделирования и параметризации педагогических тестов. – М., 2000. – 168 с.
5. О сертификации качества педагогических тестовых материалов: Приказ МО РФ от 17.04.2000 г. № 1122 // Вестник образования. – 2000. – № 11.
6. Черепанов В. С. Экспертные оценки в педагогических исследованиях. – М.: Педагогика, 1989. – 152 с.
7. Черепанов В. С., Шихова О. Ф., Шихов Ю. А., Любимова О. В. К вопросу о конструировании педагогических тестов на основе качественной технологии // Педагогические измерения в России: Материалы конф. – М., 2002. – С. 59–63.
8. Шихова О. Ф., Габдуллина Л. А. Критерии для оценки объективности педагогических контрольных материалов // Образование и наука. – 2000. – № 3 (5). – С. 82–85.
9. Экспертные методы оценки качества промышленной продукции. Основные положения. ГОСТ 23554.0–79. Организация и проведение экспертной оценки качества продукции. ГОСТ 23554. 1–79. Обработка значений экспертных оценок качества продукции. ГОСТ 23554.2–81. – М.: Изд-во стандартов, 1979; 1980; 1981.